دانشگاه علوم کشاورزی و منابع طبیعی ساری
بوم‌شناسی جنگل‌های ایران

"مقاله پژوهشی"

مقایسه عملکرد الگوریتم‌های Fuzzy C-means و K-medoids در مدل‌سازی وقوع آتش‌سوزی جنگل (مطالعه موردی: جنگل‌های سراوان، گیلان)

شقایق ذوالقدری[۱]، مهرداد قدس‌خواه دریایی[۲]، کامران نصیراحمدی[۳] و اسماعیل قجر[۴]

۱- دانشجوی دکتری گروه جنگل‌داری، دانشکده منابع طبیعی، دانشگاه گیلان (نویسنده مسوول: Shaghayegh.zolghadry@gmail.com)
۲ و ۴- دانشیار و استادیار، گروه جنگل‌داری، دانشکده منابع طبیعی، دانشگاه گیلان
۳- دکتری محیط زیست، دانشکده شیلات و محیط زیست، دانشگاه علوم کشاورزی و منابع طبیعی گرگان
تاریخ ارسال: ۹۹/۰۵/۲۶ تاریخ پذیرش: ۹۹/۰۷/۰۷
صفحه: ۱۶۳ تا ۱۷۴

چکیده

ناحیه رویشی هیرکانی (خزری) یکی از مهم‌ترین نواحی رویشی ایران محسوب شده که با توجه به قدمت آن، ارزش بوم‌سامانه‌ای بالایی دارد. از طرفی این بوم‌سامانه همه‌ساله درگیر آتش‌سوزی‌های متعدد شده و سطح قابل ملاحظه‌ای از پوشش گیاهی خود را از دست می‌دهد، لذا به‌کارگیری روش‌های علمی برای پیش‌بینی مکان‌های دارای پتانسیل خطر آتش‌سوزی در مدیریت حفاظتی جنگل‌های هیرکانی بسیار حائز اهمیت است. بسیاری از سیستم‌های دنیای واقعی از نظر تشخیص الگو مورد استفاده قرار می‌گیرند بنابراین استفاده صحیح از روش‌های یادگیری ماشین در کاربردهای عملی ضروری است. از طرفی استفاده از روش‌های مبتنی بر خوشه‌بندی با توجه به رویکرد آن در تشخیص الگو و کشف خروجی به‌عنوان یک روش موثر مورد تاکید است. هدف از انجام تحقیق حاضر بررسی توانایی و مقایسه عملکرد رویه‌های متفاوت خوشه‌بندی از دو الگوریتم مبتنی بر خوشه‌بندی Fuzzy C-Means و k-Medoids در مدل‌سازی آتش‌سوزی جنگل با تاکید بر قابلیت‌های عملکرد الگوریتم‌های موصوف است. با توجه به‌وجود آتش‌سوزی‌های دوره ای موجود از الگوریتم‌های مذکور به‌صورت ارتقاء سطح کدنویسی در نرم‌افزار متلب در راستای بهبود مطالعات در زمینه پیش‌بینی خطر حریق جنگل استفاده شد. معیارهای ورودی مدل در این مطالعه عبارتند از نقاط ثبت‌شده آتش‌سوزی، فاصله از مناطق کشاورزی، فاصله از جاده، فاصله از رودخانه، فشار هوا، بازتابش خورشید، شیب، جهت شیب، سرعت باد، درصد تراکم تاج پوشش و تیپ جنگل. نتایج به‌دست‌آمده از نقشه پیش‌بینی خطر آتش‌سوزی هر دو الگوریتم، نشان از توانایی بالای آن‌ها در پیش‌بینی مدل وقوع آتش‌سوزی دارد. همچنین بر اساس نتایج جدول ماتریس درهم‌آمیختگی مقایسه دو الگوریتم، الگوریتم FCM عملکرد بهتری نسبت به الگوریتم k-medoids در پیش‌بینی مکان‌های دارای پتانسیل خطر آتش‌سوزی از خود نشان داد. لذا استفاده از الگوریتم FCM به‌عنوان یکی از روش‌های موثر در خوشه‌بندی تفکیکی برای مطالعات آینده پیشنهاد می‌شود.

واژه‌های کلیدی: الگوریتم خوشه‌بندی، جنگل سراوان، مدل‌سازی وقوع آتش‌سوزی

مقدمه

جنگل‌ها یکی از مهم‌ترین منابع طبیعی روی زمین و تأمین‌کننده بسیاری از منابع برای انسان هستند (۲). آتش‌سوزی جنگل که جزء جدایی‌ناپذیر این بوم‌سامانه‌ها است (۵) می‌تواند منجر به نابودی این منابع شود. از طرفی از آنجاکه آتش‌سوزی‌های جنگلی سریع و به‌شدت مخرب هستند، کنترل و پایش آن فرآیندی مشکل است (۴۳). در چند سال گذشته عواملی مانند تغییرات اقلیمی و فعالیت‌های انسانی (۳) منجر به افزایش روند و تناوب آتش‌سوزی‌های جنگلی شده و نرخ آن‌ها را به مرز هشدار رسانده است (۴،۳، ۴۴،۳۷). در این میان جنگل‌های هیرکانی که یکی از ارزشمندترین جنگل‌های جهان به‌شمار می‌آیند (۱۴)، از تبعات آتش‌سوزی در امان نبوده و خسارات بسیاری را در نتیجه آتش‌سوزی متحمل شده‌اند. از این‌رو ضرورت مطالعه‌های بیشتر برای پیش‌بینی دقیق‌تر آتش‌سوزی و مدیریت آتش در چنین بوم–سامانه‌هایی بیش از پیش مشخص می‌شود. در نتیجه می‌توان در اطفای به‌موقع حریق از چنین مطالعاتی استفاده کرد (۱۲). به این منظور مدل‌های شبیه‌سازی کامپیوتری مبتنی بر مطالعات میدانی، می‌توانند درک بهتری از رفتار و اثرات آتش‌سوزی داشته باشند، در نتیجه امروزه به‌جای پژوهش‌هایی که شامل آتش‌سوزی‌های واقعی می‌شوند و نیازمند زمان و هزینه بالایی هستند از مدل‌ها و روش‌های یادگیری ماشینی استفاده می‌شود (۵۲).

پژوهش‌های بسیاری در رابطه با روش‌های مختلف مدل‌سازی آتش‌سوزی در ایران (۱، ۹، ۱۸–۲۰، ۲۴، ۲۸–۲۷) و در خارج از ایران (۱۰، ۳۶، ۴۳–۴۱) انجام شده است. با این وجود خلأهای موجود در این زمینه نشان می‌دهد هنوز نیازمند پژوهش‌های بیشتری هستیم (۵۰).

در این مطالعه در رویکردی جدید با استفاده از روش شبکه عصبی مصنوعی پرسپترون چندلایه (MLP) و روش یادگیری بدون ناظر (Unsupervised learning) با تمرکز بر الگوریتم خوشه‌بندی Fuzzy C-Means و K-Medoids به مدل‌سازی پیش‌بینی مکانی آتش‌سوزی جنگل پرداختیم. شایان ذکر است در رابطه با به‌کارگیری دو الگوریتم مذکور در مدل‌سازی آتش‌سوزی جنگل مطالعات اندکی انجام شده است. در جدیدترین پژوهش‌های انجام شده در این زمینه Giwa وBenkrid (۲۰۱۸) و خاتمی و همکاران از روش خوشه‌بندی K-Medoids برای پیش‌بینی پیکسل‌هایی که

آتش‌سوزی در آن‌ها اتفاق افتاده، استفاده کردند، که در تمامی موارد نتایج آن‌ها در استفاده از این الگوریتم رضایت بخش بود (۱۷، ۳۳-۳۲). در مطالعه‌ای دیگر جعفرزاده و همکاران (۱۳۹۸) برای ارزیابی ریسک آتش‌سوزی جنگل از الگوریتم FCM در تهیه نقشه ریسک آتش‌سوزی استفاده کردند که نقشه به‌دست آمده با مدل مذکور از دقت بالایی برخوردار بود (۲۶).

مطالعه حاضر با هدف بررسی توانایی و مقایسه الگوریتم Fuzzy C-Means و K-Medoids در مدل‌سازی آتش‌سوزی جنگل انجام شد. فرض بر این است که دو الگوریتم مذکور در پیش‌بینی حریق موفق عمل می‌کنند و عملکرد دوالگوریتم در پیش‌بینی آتش‌سوزی تفاوت چندانی ندارند. نتایج این تحقیق زمینه‌ساز پژوهش‌های بیشتر در راستای به‌کارگیری الگوریتم‌های خوشه‌بندی و بهبود مطالعات پیش‌بینی آتش‌سوزی جنگل و به‌دنبال آن عملکرد بهتر در هنگام وقوع آتش با تکیه بر مطالعات انجام شده است.

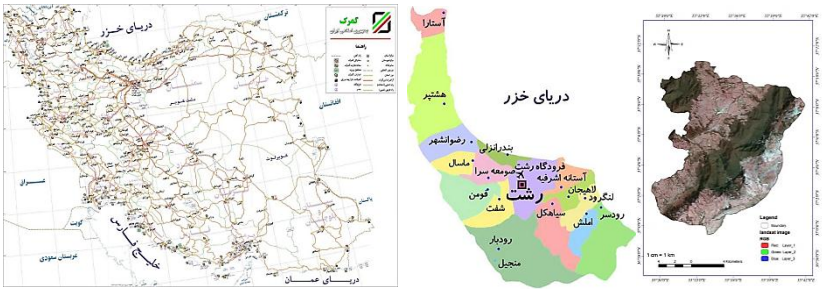
مواد و روش‌ها

منطقه مورد مطالعه

سری ۳ جنگل سراوان در قسمتی از جنگل‌های هیرکانی در استان گیلان قرار گرفته است (شکل ۱). ارتفاع این منطقه ۵۰ تا ۶۰۰ متر از سطح دریا و مساحت آن که ۸۹۳۷ هکتار است.

داده‌ها

نخست لایه‌های مورد نیاز برای مدل‌سازی شامل نقاط ثبت‌شده آتش‌سوزی، فاصله از مناطق کشاورزی، فاصله از جاده، فاصله از رودخانه، فشار هوا، بازتابش خورشید، شیب، جهت شیب، سرعت باد، درصد تراکم تاج پوشش و تیپ جنگل حاصل از تصاویر ماهواره‌ای ایجاد و معیارهای مؤثر بر آتش‌سوزی با استفاده از نرم‌افزارهای ArcGIS و ENVI برای ورود به مدل آماده‌سازی شدند و سپس این معیارها به سه بخش عوامل انسانی (فاصله از مناطق کشاورزی، جاده)، محیطی (شیب، جهت شیب، رودخانه‌ها، درصد تاج پوشش، تیپ جنگل) و اقلیمی (فشار هوا، تابش و سرعت باد) تقسیم‌بندی و نهایتاً نقشه‌های منطقه مورد مطالعه پس از تولید، برای ورود به نرم‌افزار متلب به ماتریس تبدیل شدند (۶).



شکل ۱- موقعیت سری ۳ جنگل سراوان، گیلان، ایران
Figure 1. Location of the Saravan forest in Guilan Province, Iran

روش پژوهش

در اولین مرحله داده‌کاوی، برای ورود به مدل ابتدا بین ویژگی‌های مفروض، ضریب همبستگی نمونه‌ها با یکدیگر استخراج شد. در این مطالعه از ضریب همبستگی پیرسون استفاده شده است. این ضریب ارتباط میان دو متغیر کمی را اندازه‌گیری می‌کند. هرچه مقدار ضریب بیش‌تر باشد امکان پیش‌بینی مقدار یکی از متغیرها برحسب دیگری بیش‌تر است و حدود تغییرات آن بین ۱ و ۱- است. مقدار نزدیک به ۱ و یا ۱ همبستگی زیاد و در جهت مثبت و مقدار نزدیک به ۱- و یا ۱- همبستگی زیاد و در جهت منفی میان متغیرها را نشان می‌دهد. برای محاسبه مقدار ضریب از رابطه زیر استفاده می‌شود (۲۲).

$$p(x,y)= r-(xy)\ (cov\ (x,y))/(\sigma\text{-}x\ \sigma\text{-}y) \qquad (۱)$$

در این رابطه کوواریانس، σx انحراف معیار متغیر x و σy انحراف معیار متغیر y را نشان می‌دهند.

در نهایت با استفاده از دو روش خوشه‌بندی Fuzzy C-mean و K-medoids به مدل‌سازی پیش‌بینی خطر آتش‌سوزی در منطقه سراوان با ورودی‌های مذکور پرداخته شد.

در اولین مرحله مدل‌سازی، اطلاعات ورودی نرمال شدند. در الگوریتم‌های مربوط به خوشه‌بندی در صورتی که داده‌ها بسیار ناهمگن باشند و تفاوت فاحشی در مقادیر بیشنه و کمینه آن‌ها وجود داشته باشد، معیارهای فاصله متداول و معمولاً ناکارآمد خواهند بود و بزرگی و کوچکی داده‌ها تأثیر فاحشی بر دقت جواب نهایی می‌گذارد. با توجه به اینکه معیارهای ورودی ما بر اساس تعریف‌های متغیر از اعداد مختلف با بازه‌های بسیار متفاوت تشکیل شده است، انجام پیش‌پردازش بر روی این داده‌ها امری ضروری به‌حساب می‌آید. برای از بین بردن این تفاوت‌ها با توجه به مثبت و مخالف صفر بودن این اعداد، از لگاریتم در مبنای ۱۰ استفاده شده است تا داده‌ها به‌صورت همگن‌تر در یک محدوده مشخص‌تر قرار گیرند. سپس در ادامه با استفاده از دستور polyfit، استانداردسازی تمام داده‌ها در محدوده صفر و یک انجام شد. برای تکمیل این مرحله با استفاده از دستور polyval یا نگاشت خطی (یک تابع بین دو فضای برداری که دو عملیات جمع برداری و ضرب نرده‌ای را باقی نگه می‌دارد و با عبارت عملگر خطی رابطه مستقیم دارد) بر روی داده‌ها انجام شد. تمام عملیات ذکر شده در محیط نرم‌افزار متلب

[DOR: 20.1001.1.24237140.1400.9.17.17.1]
[DOI: 10.52547/ifej.9.17.163]

انجام پذیرفته است (۲۱). ضمناً داده‌های کیفی همگی با رویه transformation به داده‌های کمی تبدیل شدند و سپس مورد استفاده قرار گرفتند.

مرحله بعدی استفاده از روش‌های یادگیری خوشه‌بندی است. این روش یکی از شاخه‌های تحلیل آماری چندمتغیره و یادگیری بدون ناظر در شبکه عصبی مصنوعی است یعنی یادگیری در این نوع شبکه‌ها به جواب مطلوب و مقدار واقعی پاسخ از پیش مشخص، دسترسی ندارد و در بین ورودی‌ها جواب مطلوب را پیدا می‌کند و در حقیقت این روش، یادگیری از طریق داده‌ها و مشاهدات است و قادر است اطلاعات مفید را از میانگین حجم بالای داده‌ها استخراج کند. در خوشه‌بندی جامعه به تعدادی زیرجامعه به‌نام خوشه تقسیم می‌شود. در این خوشه‌ها نمونه‌هایی که به یکدیگر شبیه هستند و با نمونه‌های خوشه‌های دیگر شباهتی ندارند (۳۷)، دسته‌بندی می‌شوند (۳۰-۳۱). دسته‌بندی این خوشه‌ها بر اساس روابط بین آن‌ها صورت می‌گیرد (۲۱).

در خوشه‌بندی ابتدا ماتریس داده‌ها تهیه و ارائه و در مرحله بعدی استاندارد می‌شود (طبق آنچه در پاراگراف بالا توضیح داده شد). سپس ماتریس مجاورت یا مشابهت محاسبه و روش خوشه‌بندی اجرا می‌شود. در نهایت معیارهای اعتبار محاسبه می‌شوند (۳۱) که تمامی موارد موصوف در محیط نرم‌افزار متلب انجام شده است.

اولین خوشه‌بندی مورد استفاده در این پژوهش یعنی k-medoids بهبود یافته الگوریتم k-means است و بسیار شبیه به این الگوریتم عمل می‌کند، با این تفاوت که در الگوریتم k-medoids به‌جای استفاده از میانگین، از خود نمونه‌ها برای مرکز ثقل و نمایندگی خوشه‌ها استفاده می‌شود . قابل ذکر است در این الگوریتم هدف کمینه کردن مجموع اختلافات میان نقاط موجود در یک خوشه‌بندی و نقطه مرکز خوشه است. در صورتی‌که در الگوریتم k-means هدف حداقل‌کردن مربعات خطا است. در این الگوریتم مرکز خوشه‌ها به‌جای centroid، medoid است. هر medoid مرکزی‌ترین داده یک خوشه است. در کل از آن‌جایی که این الگوریتم حساسیت کمی نسبت به داده‌های خارج از محدوده دارد و همچنین با توجه به ثبت نقاط فعلی در منطقه مورد مطالعه که آتش‌سوزی در آنها در سنوات قبل اتفاق افتاد، استفاده از این الگوریتم نسبت به الگوریتم k-means در این مطالعه دارای ارجحیت بوده و استفاده از آن بیشتر از الگوریتم k-means توصیه می‌شود (۳۱،۲۵). برای اجرای الگوریتم k-medoids ابتدا مقدار k داده به‌صورت تصادفی به‌عنوان نماینده‌های اولیه k خوشه انتخاب می‌شوند و برای نمونه نزدیک‌ترین نماینده خوشه پیدا می‌شود. پس از تشکیل ماتریس تشابه (n-k) نمونه در یکی از این k خوشه قرار می‌گیرند. می‌توان به‌جای تشکیل ماتریس تشابه، فاصله هر یک از نمونه‌های باقی‌مانده را با k نمونه اولیه محاسبه کرد. سپس برای بررسی کیفیت و مناسب بودن خوشه‌های به‌دست‌آمده، یک نمونه از داده‌ها (غیر مدوید یا غیر از داده‌های مورد انتخاب شده برای مرکز خوشه) با یکی از k نمونه نماینده (مدوید یا عنصر مرکزی خوشه) جایگزین می‌شوند. هزینه حاصل از این جایگزینی محاسبه و چنانچه منفی بود این جابه‌جایی صورت می‌گیرد. این مرحله تا زمانی که مقدار مراکز خوشه در دو گام متوالی ثابت بماند، تکرار می‌شود (۳۱). در خوشه‌بندی با الگوریتم k-medoids الگوریتم‌های متفاوتی وجود دارد. یکی از استانداردترین و مرسوم‌ترین همچنین قدرتمندترین روش‌ها در این الگوریتم، PAM است. البته این الگوریتم به‌دلیل نیازمندی به تکرار زیاد مناسب مجموعه داده‌های حجیم نیست (۳۱).

دو الگوریتم CLARA و CLARANS نیز در روش k-medoids وجود دارند. برای مجموعه داده‌های بزرگ از الگوریتم CLARA استفاده می‌شود. در این الگوریتم نمونه‌های تصادفی از مجموعه داده انتخاب شده و الگوریتم PAM برای آن‌ها به‌کار گرفته می‌شود و بهترین خوشه‌بندی را به‌عنوان خروجی ایجاد می‌کند. در پایان، تخصیص باقی اجزای پایگاه داده بر اساس خروجی ایجاد شده به نزدیک‌ترین خوشه انجام می‌شود. دیگر الگوریتم مورد استفاده در k-medoids الگوریتم CLARANS است که بر مبنای نشان دادن داده‌های دور از مرکز است (۵۲). با توجه به اینکه مجموعه داده‌های ما خیلی بزرگ نیستند در این مطالعه از الگوریتم PAM استفاده شد (۲۵).

معمول‌ترین معیار محاسبه فاصله داده‌ها در الگوریتم k-medoids فاصله اقلیدسی است. در این روش فاصله داده‌ها با مرکز ارزیابی و با مرکز متناظر خود محاسبه و در متغیر مشخصی ذخیره شده و در مرحله بعدی فضای احتمال بالای حادثه با فرض حداکثر ۱۵ ٪ از بیشترین فاصله نقطه حادثه‌خیز از مرکز خوشه متناظر با خود محاسبه می‌شود (۱۱، ۳۴،۲۹،۲۵). در این تحقیق نیز از همین روش در نرم‌افزار متلب استفاده شده است.

دومین الگوریتم مورد استفاده در پژوهش حاضر Fuzzy c-means است. در این الگوریتم هر نقطه می‌تواند با درجات عضویت مختلف به بیش از یک خوشه تعلق گیرد. در حقیقت عضویت در آن ماهیت فازی دارد و در بازه ۱-۰ قرار می‌گیرد. این نوع خوشه‌بندی وقتی استفاده می‌شود که نمی‌توان هر داده را دقیقاً به یک خوشه اختصاص داد زیرا برخی داده‌ها بین خوشه‌ها قرار می‌گیرند (۷). هدف از این خوشه‌بندی، گروه‌بندی n بردار p بعدی در c خوشه است (۳۹). میزان عضویت هر شیء در FCM به هر خوشه از ماتریس عضویت مشخص می‌شود (رابطه ۲).

(رابطه ۲)

$$U = [u_{ij}]c.n = (u_1, u_2, \dots u_n)$$

در این رابطه c تعداد خوشه‌ها و n تعداد اشیاء U ماتریس عضویت است. هدف الگوریتم خوشه‌بندی فازی، حداقل کردن تابع هدف یا هزینه برای یک مجموعه داده است. در تحقیق حاضر ۵۷ فقره آتش‌سوزی (n) و ۵ خوشه (c) در نظر گرفته شده است.

در رابطه (۳)، dij فاصله بین داده‌ها و X مرکز خوشه است. $m \in (1, \infty)$ میزان فازی بودن است.

[DOR: 20.1001.1.24237140.1400.9.17.17.1]
[DOI: 10.52547/ifej.9.17.163]

(رابطه ۳)

$$J_f\left(X, U_f, C\right) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m}\, d_{ij}^{2}$$

اگر m به سمت ۱ میل کند خوشه‌بندی سخت‌تر خواهد شد و اگر m به سمت بی‌نهایت میل کند، خوشه‌بندی فازی‌تر خواهد بود. این الگوریتم به تعداد اولیه خوشه‌ها و مکان اولیه مراکز خوشه‌ها وابسته است (۴۸). در روش FCM تصمیم‌گیری بر اساس بیشینه تعلق به یک مرکز ارزیابی می‌شود (۳۹).

صحت‌سنجی مدل

برای ارزیابی قدرت پیش‌بینی و عملکرد مدل، جذر میانگین مربعات خطا (RMSE) و مقدار ضریب تبیین (R^2) محاسبه شد (۴۶). جذر میانگین مربعات خطا و مقدار ضریب تبیین، تفاوت میان مقادیر پیش‌بینی‌شده و مقادیر واقعی مشاهداتی را با استفاده از یک مدل اندازه‌گیری می‌کنند. در حقیقت ضریب تبیین مشخص می‌کند چند درصد تغییرات متغیر وابسته متأثر از متغیر مستقل مربوطه است و جذر میانگین مربعات خطا، خطاهای پیش‌بینی یک مجموعه داده مقایسه می‌کند. هرچه مقدار RMSE و R^2 بیشتر باشد، داده‌ها تطبیق بهتری با یکدیگر دارند. مقدار RMSE و R^2 از صورت محاسبه می‌شود (۳۵،۸).

(رابطه ۴)

$$R^2 = \sum_{i=1}^{n}(O_i - \bar{O}_i)(P_i - \bar{P}_i)^2 / (P_i - \bar{P}_i)^2\,(P_i - \bar{P}_i)^2$$

(رابطه ۵)

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n}(O_i - P_i)/n}$$

در این رابطه Oi داده‌های مشاهداتی، Pi داده‌های شبیه‌سازی‌شده، P شبیه‌سازی‌شده به‌وسیله مدل و n تعداد داده‌ها است. همچنین برای برآورد میزان خطای مدل از MSE یا خطای میانگین مربعات (Mean Squared Error) استفاده شد. در حقیقت خطای میانگین مربعات تفاوت بین مقادیر تخمینی و آنچه تخمین زده شده، است و مقدار آن از رابطه زیر محاسبه می‌شود (۴۰):

(رابطه ۶)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

در این رابطه عمل میانگین‌گیری با $\sum_{i=1}^{n}$ انجام می‌شود و $(y_i - \hat{y}_i)^2$ محاسبه مقدار مربع خطای هر داده را به عهده دارد (۴۶).

در نهایت برای مقایسه عملکرد دو الگوریتم از ماتریس درهم‌ریختگی (Confusion Matrix) برای ارزیابی نحوه عملکرد دو الگوریتم مورد استفاده در روش خوشه‌بندی به‌منظور حصول بالاترین دقت و صحت در انتخاب خوشه‌ها استفاده شده است. در حقیقت این ماتریس جدولی است که عملکرد مدل طبقه‌بندی‌شده را توصیف می‌کند. این جدول اطلاعاتی در مورد طبقه‌بندی واقعی و پیش‌بینی صورت گرفته توسط طبقه‌بندی را ارائه می‌دهد که با استفاده از این اطلاعات عملکرد طبقه‌بندی ارزیابی می‌شود (۱۶).

نتایج و بحث

ضریب همبستگی پیرسون

جدول تحلیل ضریب همبستگی این امکان را به طراح و تحلیل‌کننده می‌دهد که در صورت لزوم ویژگی‌هایی را که دارای شباهت بسیار بالا با یکدیگر هستند از فرآیند تصمیم‌گیری حذف کند و فقط از یکی از آن‌ها استفاده کند (جدول ۱). در جدول (۱) قطر اصلی با عدد ۱ مشخص شده است. در نهایت این جدول نشان می‌دهد داده‌ها از نظر ضریب همبستگی به‌عنوان ورودی شباهت چندانی به‌هم نداشته و استفاده از آنها برای فرایند مدل‌سازی مطلوب است.

جدول ۱- جدول تحلیل ضریب همبستگی پیرسون

Table 1. Pearson correlation coefficient analysis table

پارامترهای مورد بررسی	سرعت باد	تیپ جنگل	تابش خورشید	درجه شیب	فاصله از جاده	فاصله از رودخانه	فاصله از مناطق کشاورزی	فشار هوا	جهت شیب	تاج پوشش جنگلی
تاج پوشش جنگلی	-۰/۱۱۷۹	۰/۸۷۸۷	-۰/۳۳۹۲	-۰/۰۶۷۹	۰/۵۴۲۴	۰/۱۷۷۵	۰/۰۵۰۶	۰/۵۱۳۹	۰/۱۴۱۶	۱/۰۰۰
جهت شیب	-۰/۰۳۵۰	۰/۱۵۲۶	-۰/۰۱۰۱	-۰/۰۸۵۹	۰/۰۵۸۸	-۰/۰۳۰۶	۰/۰۱۳۵	-۰/۰۲۵۱	۱/۰۰۰	۰/۱۴۱۶
فشار هوا	-۰/۲۵۹۸	۰/۶۱۲۱	-۰/۳۹۲۹	۰/۰۲۰۷	۰/۶۱۴۶	۰/۰۳۵۶	۰/۱۷۱۱	۱/۰۰۰	-۰/۰۲۵۱	۰/۵۱۳۹
فاصله از مناطق کشاورزی	۰/۱۷۹۵	۰/۰۶۴۳	۰/۰۸۱۸	-۰/۰۷۹۱	۰/۱۲۴۶	-۰/۰۰۲۵	۱/۰۰۰	۰/۱۷۱۱	۰/۰۱۳۵	۰/۰۵۰۶
فاصله از رودخانه	۰/۰۶۳۷	-۰/۰۷۴۲	۰/۱۲۶۶	-۰/۳۸۶۵	-۰/۰۵۷۲	۱/۰۰۰	-۰/۰۰۲۵	۰/۰۳۵۳	-۰/۰۳۰۶	۰/۱۷۷۵
فاصله از جاده	-۰/۲۶۱۳	۰/۷۲۳۴	-۰/۳۲۶۹	۰/۰۶۴۴	۱/۰۰۰	-۰/۰۵۷۲	۰/۱۲۴۶	۰/۶۱۴۶	۰/۰۵۸۸	۰/۵۴۲۴
درجه شیب	۰/۰۷۲۳	۰/۰۹۴۱	-۰/۴۸۲۸	۱/۰۰۰	۰/۰۶۴۴	-۰/۳۸۶۵	-۰/۰۷۹۱	۰/۰۲۰۷	-۰/۰۸۵۹	-۰/۰۶۷۹
تابش خورشید	۰/۲۰۵۷	-۰/۴۵۷۸	۱/۰۰۰	-۰/۴۸۲۸	-۰/۲۳۶۹	۰/۱۲۶۶	۰/۰۸۱۸	-۰/۳۹۲۹	-۰/۰۱۰۱	-./۳۳۹۲
تیپ جنگل	-۰/۲۷۷۹	۱/۰۰۰	-۰/۴۵۷۸	۰/۰۹۴۱	۰/۷۲۳۴	-۰/۰۷۴۲	۰/۰۶۴۳	۰/۶۱۲۱	۰/۱۵۲۶	۰/۸۷۸۷
سرعت باد	۱/۰۰۰	-۰/۲۷۷۹	۰/۲۰۵۷	-۰/۰۷۲۳	-۰/۲۶۱۳	۰/۰۶۳۷	۰/۱۷۹۵	-۰/۲۵۹۸	-۰/۰۳۵۰	-۰/۱۱۷۹

رویه کلی تخصیص نقاط دارای پتانسیل خطر آتش سوزی در الگوریتم K-Medoids و FCM

مراحل انجام هر دو الگوریتم تقریباً به یکدیگر شبیه است و تفاوت آن‌ها در نحوه تخصیص مرکز داده‌ها است. یعنی در الگوریتم K-medoids مرکز داده‌ها medoid و در FCM مرکز داده‌ها centroid است و به‌صورت فازی ارزیابی می‌شود (۵۲). در هر دو نوع الگوریتم تعداد ۵ خوشه و ۵ مرکز خوشه داریم. سطح آستانه مورد نظر در هر دو خوشه ۱ کیلومتر در

[DOR: 20.1001.1.24237140.1400.9.17.17.1]
[DOI: 10.52547/ifej.9.17.163]

نظر گرفته شد (جداول ۳ و ۴). در دو الگوریتم مورد استفاده داده‌ها از نظر فاصله با مرکز ارزیابی شده و فاصله هر داده با مرکز متناظر خودش محاسبه شده است. با توجه به اینکه فضای احتمال بالای حادثه با فرض بالای حداکثر ۱۵٪ از بیشترین فاصله نقطه حادثه‌خیز از مرکز خوشه متناظر خودش محاسبه شده است با افزایش و کاهش از این فاصله احتمال رخداد آتش به‌ترتیب کمتر و بیشتر خواهد شد.

جدول ۲- خوشه‌بندی پارامترهای مورد مطالعه به روش K-Medoids

Table 2. Clustering of the studied parameters by K-Medoids

مرکز	سطح آستانه	سرعت باد	تیپ جنگل	تابش خورشید	شیب	فاصله از جاده	فاصله از رودخانه	فاصله از مناطق کشاورزی	فشار هوا	جهت شیب	کلاس تاج پوشش
۱	۰/۷۱۳	۰/۷۱۶	۰/۰۰۰	۰/۳۴۱	۰/۷۸۶	۰/۶۸۴	۰/۵۱۷	۰/۶۲۴	۰/۶۹۲	۰/۵۹۳	۰/۳۲۹
۲	۱/۳۲۴	۰/۵۰۱	۱/۰۰۰	۰/۹۲۳	۰/۷۰۵	۰/۵۸۷	۰/۸۲۹	۰/۶۸۵	۰/۷۰۷	۰/۰۵۸۹	۰/۷۲۲
۳	۱/۸۷	۰/۶۵۸	۰/۰۰۰	۰/۱۴۴	۰/۷۸۳	۰/۶۸۴	۰/۵۳۳	۰/۶۳۹	۰/۶۴۷	۰/۳۲۹	۰/۱۲۸
۴	۰/۸۱۲	۰/۷۰۰	۱/۰۰۰	۰/۰۵۹	۰/۸۰۹	۰/۶۸۲	۰/۵۳۲	۰/۶۸۱	۰/۶۲۷	۰/۲۲۸	۰/۰۵۲
۵	۰/۶۳۰	۰/۴۹۹	۰/۳۳۳	۰/۹۶۱	۰/۴۲۵	۰/۷۷۷	۰/۸۸۸	۰/۵۹۱	۰/۶۷۳	۰/۶۸۸	۰/۶۶۰

جدول ۳- خوشه‌بندی پارامترهای مورد مطالعه به روش FCM

Table 3. Clustering of the studied parameters by FCM

مرکز	سطح آستانه	سرعت باد	تیپ جنگل	تابش خورشید	شیب	فاصله از جاده	فاصله از رودخانه	فاصله از مناطق کشاورزی	فشار هوا	جهت شیب	کلاس تاج پوشش
۱	۰/۸۰۲	۰/۲۲۹	۰/۰۰۰	۰/۹۵۲	۰/۸۰۷	۰/۲۸۱	۰/۴۴۳	۰/۴۶۰	۰/۳۲۹	۰/۴۷۲	۰/۰۰۰
۲	۰/۷۳۳	۰/۳۰۷	۱/۰۰۰	۰/۸۱۳	۰/۴۶۹	۰/۹۹۰	۰/۷۹۸	۰/۷۰۳	۰/۶۱۷	۰/۵۹۸	۰/۶۶۷
۳	۰/۸۶۴	۰/۷۷۶	۰/۰۰۰	۰/۷۰۳	۰/۶۵۱	۰/۵۰۸	۰/۶۹۲	۰/۶۴۷	۰/۱۸۹	۰/۷۶۷	۰/۸۰۲
۴	۰/۶۴۱	۰/۶۲۳	۱/۰۰۰	۰/۲۵۷	۰/۸۲۳	۰/۸۸۲	۰/۶۲۰	۰/۶۳۰	۰/۷۱۹	۰/۴۸۰	۰/۶۶۷
۵	۱/۲۹۹	۰/۸۰۲	۰/۳۳۳	۰/۸۱۶	۰/۵۱۵	۰/۴۳۲	۰/۵۴۰	۰/۷۰۰	۰/۳۰۹	۰/۵۹۸	۰/۳۳۳

مدل

با توجه به نتایج مشخص می‌شود میزان RMSE، R^2 و MSE برای مدل شبکه عصبی مورد استفاده در این تحقیق به‌ترتیب برابر است با ۰/۲۸۶۱ و ۹۹/۳۸ و ۰/۰۸۱۹، که نشان‌دهنده قابل اطمینان بودن مدل است.

نقشه نهایی حاصل از خوشه‌بندی با FCM و K- Medoids

در نقشه حاصل قسمت‌های سفید مربوط به مناطقی است که احتمال وقوع آتش در آن بالا و نقاط مشخص شده با رنگ قرمز مکان‌هایی با احتمال کمتر خطر آتش‌سوزی است (شکل ۲ و ۳).

مقایسه دو الگوریتم

براساس نتایج حاصل از جدول تحلیل ماتریس درهم‌ریختگی طبق دو الگوریتم مورد استفاده ۵ طبقه احتمال آتش‌سوزی داریم. طبقه ۱ که فاصله آن از نزدیک‌ترین مرکز خوشه آتش‌سوزی کمتر از ۱ کیلومتر است طبقه لکه‌های داغ (Hot spot) نام دارد. طبقه ۲، طبقه پرخطر درجه یک است که در آن فاصله از نزدیک‌ترین مرکز خوشه ۵-۱ کیلومتر است. طبقه سوم مربوط فاصله ۱۰-۵ کیلومتری از مرکز خوشه است و طبقه پرخطر درجه دوم نامیده می‌شود. طبقه دیگر طبقه کم‌خطر درجه ۱ با فاصله ۵۰-۱۰ کیلومتری از مرکز خوشه است و طبقه آخر که فاصله ۲۵۵-۵۰ کیلومتری از مرکز خوشه را به‌خود اختصاص می‌دهد مربوط به طبقه کم‌خطر درجه ۲ است (شکل ۴).

به‌طور کل در جدول نتایج حاصل از ماتریس درهم‌ریختگی سه جدول وجود دارد که به‌ترتیب شامل جدول پیکسل‌های محدوده مورد مطالعه (جدول بزرگ)، جدول بررسی سطوح پیکسلی محدوده مطالعه با پیش فرض بودن الگوریتم FCM (جدول سمت راست) و جدول بررسی سطوح پیکسلی محدوده مورد مطالعه به‌صورت ستونی با اولویت الگوریتم K-medoids (جدول پایینی) می‌باشد. در تحلیل سطری ستون‌های آبی رنگ مربوط به درصد تشابه پیش‌بینی هر دو الگوریتم است بر همین اساس ۳۸/۷٪ پیکسل‌ها در هر دو الگوریتم در طبقه هات اسپات، ۷۹/۱٪ در طبقه پرخطر درجه اول، ۶۳/۹ در طبقه پرخطر درجه دوم، ۹۶/۵، در طبقه کم‌خطر درجه اول و ۱۰۰٪ در طبقه کم‌خطر درجه دوم قرار دارند. همچنین سلول‌های نارنجی رنگ جدول موصوف نشان از پیش‌بینی مواردی از آتش‌سوزی توسط الگوریتم FCM دارد که توسط الگوریتم K-medoids مورد تایید قرار نگرفته است و در حقیقت نقطه مورد اختلاف دو الگوریتم در فرایند مدل‌سازی است. بر اساس نتایج این جدول ۶۱/۳٪ توسط الگوریتم FCM در طبقه هات اسپات، ۲۰/۹٪ در طبقه پرخطر درجه اول، ۳۶/۱ در طبقه پرخطر درجه دوم، ۹/۵٪ در طبقه کم‌خطر درجه اول قرار می‌گیرند.

در تحلیل ستونی که با اولویت الگوریتم K-medoids است، ۴۹/۲٪ توسط هر دو الگوریتم به‌عنوان طبقه هات اسپات، ۷۲/۱ ٪ طبقه پرخطر درجه اول، ۳۲/۱ طبقه پرخطر درجه دوم، ۹۶/۵ ٪ طبقه کم خطر درجه اول و ۱۰۰٪ طبقه پرخطر درجه دوم شناسایی شده اند که با رنگ آبی نشان داده شده است.

در این جدول سلول‌های نارنجی مربوط به موارد پیش‌بینی خطر آتش‌سوزی است که تنها توسط الگوریتم K-medoids در نظر گرفته شده است که مورد تایید الگوریتم FCM نیست. نتایج حاصل از آن نشان داد که تعداد ۵۰/۸٪ پیکسل‌ها در

[DOR: 20.1001.1.24237140.1400.9.17.17.1]
[DOI: 10.52547/ifej.9.17.163]

طبقه هات اسپات، ۲۷/۹٪ در طبقه پرخطر درجه اول، ۶۷/۹ در طبقه پرخطر درجه دوم و ۳/۵ ٪ در طبقه کم‌خطر درجه اول قرار دارند. به‌طور کل افزایش سطح اشتراک نتایج دو الگوریتم در طبقات بالاتر چه در تحلیل سطری و ستونی محدوده مورد مطالعه، تاییدکننده کاهش حساسیت این الگوریتم‌ها با فاصله گرفتن از مراکز خوشه می باشد. با توجه به اینکه مجموع اشتراک دو الگوریتم با تاکید بر مبنا بودن الگوریتم فازی ۳۷۲/۲ ٪ و مجموع اشتراک بر مبنای الگوریتم k-medoids برابر با ۳۴۹ ٪ است، مشخص می‌شود الگوریتم فازی در پیش‌بینی مکان‌های دارای پتانسیل خطر آتش سوزی با اختلاف جزئی بهتر از الگوریتم k-medoids عمل نموده است که در سایر تحقیقات نیز اثبات شده است.

نتایج تحقیقات خاتمی و همکاران در سال ۲۰۱۷ در شناسایی آتش با استفاده از پردازش تصویر با به‌کارگیری الگوریتم k-medoids در تحلیل خوشه‌بندی نشان داد که این الگوریتم به تعداد تکرار زیادی بیشتری نیاز دارد و به‌همین دلیل عملکرد آن نسبت به FCM پایین‌تر است (۳۰). در مورد کار ما نیز بر اساس نتایج (شکل ۲) مشخص می‌شود، FCM عملکرد بهتری نسبت به k-medoids دارد. همچنین بر اساس نتایج مطالعات انجام‌شده توسط Esakar و همکاران در به‌کارگیری الگوریتم FCM در تحلیل خوشه‌بندی، با به‌کارگیری این الگوریتم در داده‌های خوشه‌بندی در آموزش شبکه عصبی ارتباط میان داده‌های ورودی و خروجی بهتر می‌شود و در نتیجه احتمال دقت پیش‌بینی بالا می‌رود (۱۵، ۴۸).

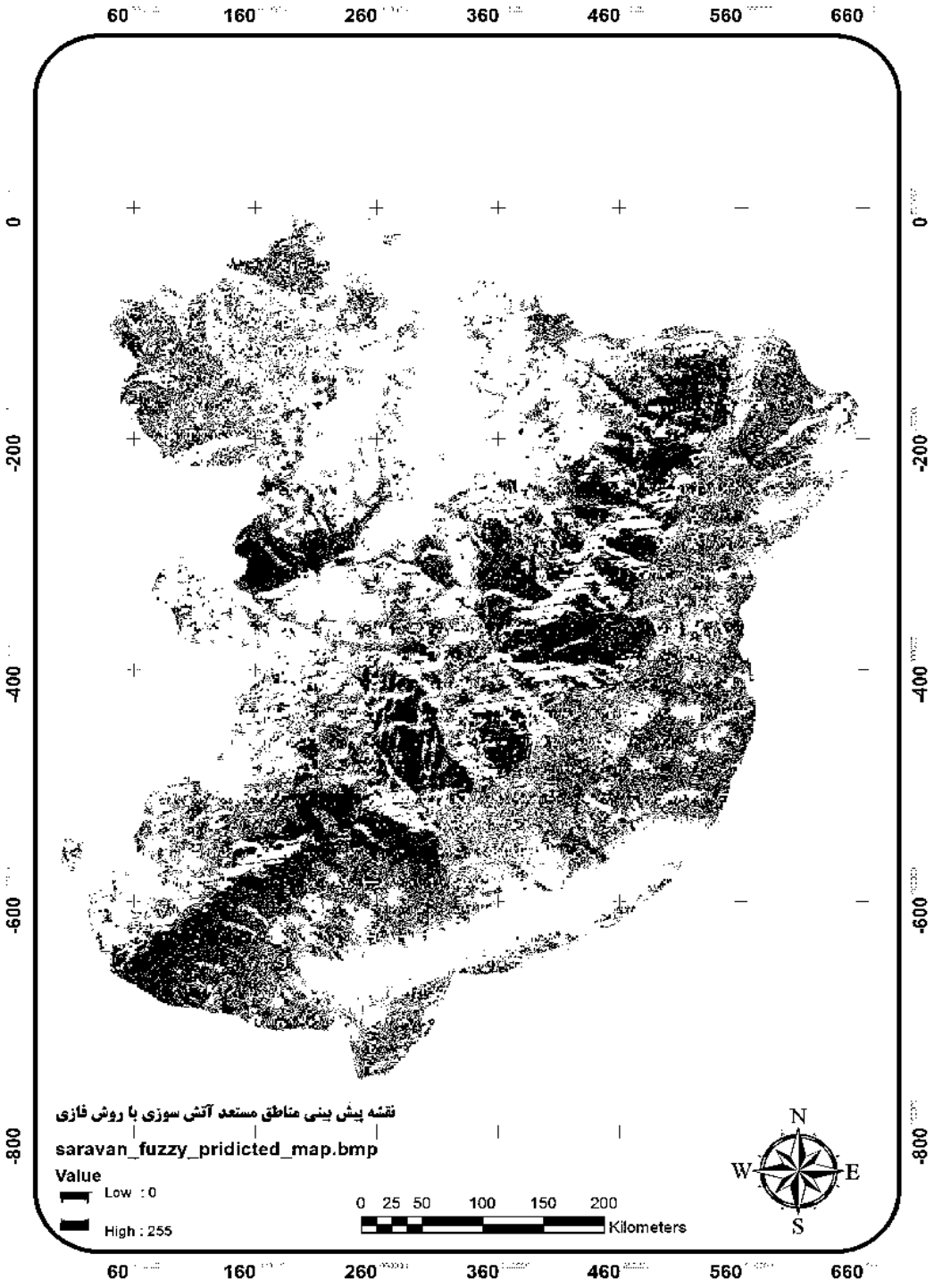
در بعضی مطالعات این نتایج متفاوت است. در تحقیقی دیگر Velmurugan (۲۰۱۱) نشان داد k-medoids عملکرد بهتری نسبت به FCM در خوشه‌بندی داده‌های مخابراتی دارد و علت آن این است که FCM نسبت به k-medoids زمان بیشتری برای اجرای الگوریتم نیاز دارد (۴۷). مطالعات دیگری که توسط همین محقق در رابطه با مقایسه دو الگوریتم FCM و k-medoids در سال ۲۰۱۲ انجام شده نشان داده است که کی از مهم‌ترین مسائل در اجرای الگوریتم k-medoids زمان مورد نیاز برای جایگزینی عناصر مدوید یا همان عناصر مرکزی خوشه است و با افزایش k تعداد خوشه‌ها افزایش می‌یابد و در نتیجه زمان اجرای الگوریتم نسبت به FCM با همان تعداد داده، افزایش می‌یابد که منجر به کاهش عملکرد k-medoids نسبت به FCM می‌شود. بر اساس نتایج او انتخاب الگوریتم خوشه‌بندی بستگی به هدف همچنین نوع داده‌های موجود دارد. به‌طور کل با افزایش تعداد خوشه‌ها و تعداد نقاط انتخابی زمان اجری الگوریتم بالا می‌رود و عملکرد الگوریتم کاهش می‌یابد (۴۸). نکته آخر در مورد جدول تحلیل ماتریس درهم‌ریختگی در این پژوهش این است که مقادیر ۵ کلاس طبقه‌بندی فاصله از خطر در کد نوشته شده در نرم‌افزار متلب قابل تغییر بوده و با توجه حساسیت مورد نیاز برای تصمیم‌گیری‌های مدیریتی قابل تغییر (افزایش یا کاهش) می‌باشد.

با توجه به آتش‌سوزی‌هایی که هر ساله رخ می‌دهد توسعه یه سیستم شناسایی آتش بسیار ضروری است.

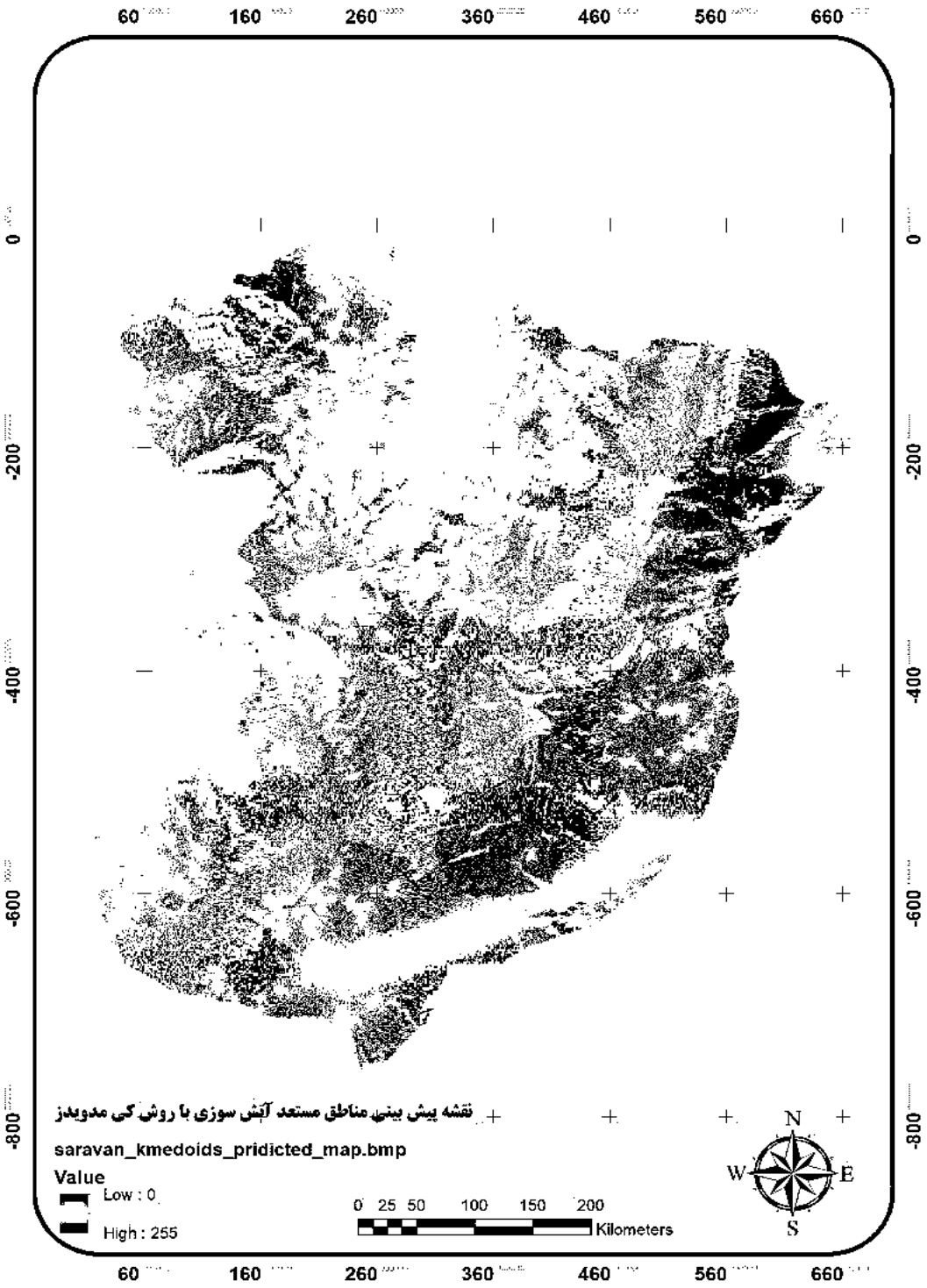
تشکر و قدردانی

بدینوسیله از زحمات دکتر عبدالرضا علوی قره باغ که با نظرات ارزشمندشان به ما در بهبود نتایج مطالعه حاضر یاری رساندند کمال تشکر را داریم.

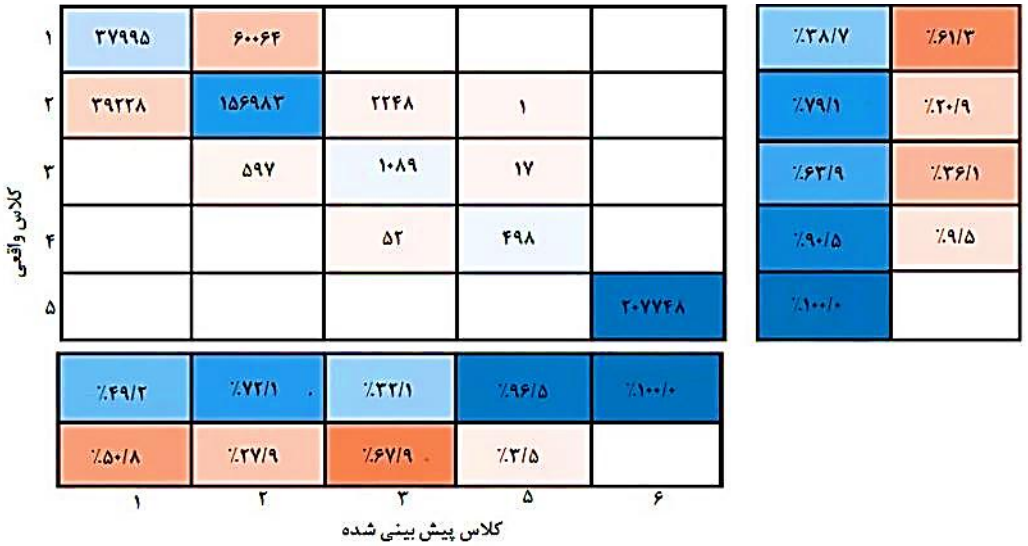
[DOR: 20.1001.1.24237140.1400.9.17.17.1]
[DOI: 10.52547/ifej.9.17.163]



شکل ۲- نقشه نهایی پیش‌بینی مناطق مستعد خطر آتش‌سوزی با Fuzzy C-means
Figure 2. Prediction map of fire risk areas with Fuzzy C-means



شکل ۳- نقشه نهایی پیش‌بینی مناطق مستعد خطر آتش‌سوزی با K-medoids
Figure 3. Prediction map of fire risk areas with K-medoids



شکل ۴- جدول تحلیل ماتریس درهم‌ریختگی
Figure 4. Confusion matrix analysis table

منابع

1. Adab, H., K. Kanniah and K. Solaimani. 2013. Modeling forest fire risk in the northeast of Iran using remote sensing and GIS techniques. Natural Hazards, 65(3): 1723-1743.
2. Agnoletti, M. and S. Anderson. 2000. Forest history: International studies on socio-economic and forest ecosystem change: report no. 2 of the IUFRO Task Force on Environmental Change / edited by M. Agnoletti and S. Anderson. Wallingford. CABI Pub in association with IUFRO. https://books.google.com/books?id=0znQhwyb6PAC.
3. Argañaraz, J.P., G.G. Pizarro, M. Zak, M.A. Landi and L.M. Bellis. 2015. Human and biophysical drivers of fires in Semiarid Chaco mountains of Central Argentina. Science of the Total Environment, 520: 1-12.
4. Arpaci, A., B. Malowerschnig, O. Sass and H. Vacik. 2014. Using multivariate data mining techniques for estimating fire susceptibility of Tyrolean forests. Applied Geography, 53: 258-70.
5. Baheri, H., M. Ghodskhah Daryaei and H. Pourbabaei. 2017. Long- Term Effect of Fire on Woody Species Composition and their Natura Regeneration in Hyrcanian Forests, (Case Study: Lesakouti Forest of Tonekabon, Mazandaran Province). Ecology of Iranian Forest, 5: 37-46 (In Persian).
6. Berninger, F. 1994. Simulated irradiance and temperature estimates as a possible source of bias in the simulation of photosynthesis. Agricultural and Forest Meteorology, 71(1-2): 19-32.
7. Bezdek, J.C., R. Ehrlich and W. Full. 1984. FCM: The Fuzzy C-Means Clustering Algorithm. Computers and Geosciences, 10(2): 191-203.
8. Chai, T. and R. Draxler. 2014. Root mean square error (RMSE) or mean absolute error (MAE)? Geosci. Model Dev. Discuss, 7(1): 1525-34.
9. Darvishi, L., M. Ghodskhah Daryaei and V. Gholami. 2013. A regional model for forest fire hazard zonation in forests of Dorud city (Case Study: Babahar region). Iranian Journal of Forest and Range Protection Research, 11(1): 10-20 (In Persian).
10. Dubey, V., P. Kumar and N. Chauhan. 2018. Forest Fire Detection System Using IoT and Artificial Neural Network. In: Bhattacharyya S, Hassanien AE, Gupta D, Khanna A, Pan I, editors. International conference on innovative computing and communications: Proceedings of ICICC. Volume 1 / edited by Siddhartha Bhattacharyya, Aboul Ella Hassanien, Deepak Gupta, Ashish Khanna, Indrajit Pan. Singapore: Springer, 55: 323-37.
11. Dunn, J.C. 1973. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. Journal of Cybernetics, 3(3): 32-57.
12. Eastaugh, C.S. and H. Hasenauer. 2014. Deriving forest fire ignition risk with biogeochemical process modelling. Environmental Modelling and Software, 55: 132-42.
13. Eskandari, S. and E. Chuvieco. 2015. Fire danger assessment in Iran based on geospatial information. International Journal of Applied Earth Observation and Geoinformation, 42: 57-64.
14. Eskandari, S. and J.R. Miesel. 2016. Comparison of the fuzzy AHP method, the spatial correlation method, and the Dong model to predict the fire high-risk areas in Hyrcanian forests of Iran. Geomatics, Natural Hazards and Risk, 8(2): 933-49.

15. Esakar, S. and M. Chaudhari. 2013. A Review of Clustering Algorithms. www.ijcst.com.
16. Fawcett, T. 2006. An introduction to ROC analysis. Pattern Recognition Letters, 27(8): 861-74.
17. Giwa, O. and A. Benkrid. 2018. Fire detection in a still image using colour information.
18. Computer Science, Engineering, 3(3) 2018.
19. Ghodskhah Daryayi, M., M.N. Adel, M.S. Pashaki and J.S. Kuhestani. 2013. Effect of repeated fire on understory plant species diversity in Saravan forests, northern Iran. Folia Forestalia Polonica, Seria A Forestry, 55(3) (In Persian).
20. Goleiji, E., S.M. Hoseini, N. Khorasani and S.M. Monavari. 2018. Forest fire risk assessment using WLC and ANP (Case study: 33 and 34 watersheds north of Iran). Journal of Natural environment hazards, 7(15): 107-24.
21. Goleiji, E., S.M. Hosseini, N. Khorasani and S.M. Monavari. 2017. Forest fire risk assessment-an integrated approach based on multicriteria evaluation. Environmental monitoring and assessment, 189(12): 612.
22. Han, J., M. Kamber and A. Tung. 2001. Spatial Clustering Methods in Data Mining: A Survey. In H. J. Miller & J. Han (eds.), Geographic Data Mining and Knowledge Discovery, Research Monographs in GIS: Taylor and Francis.486 pp.
23. Hastie, T., R. Tibshirani, J. Friedman and J. Franklin. 2004. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Math. Intell, 27: 83-5.
24. Hedayati, N, S. Ebrahimi and H. Joneidi. 2019. Fire risk assessment of Kurdistan province natural areas using statistical index method. Journal of natural environment, 403-16 (In Persian).
25. Hunt, R.J. 1986. Percent agreement, Pearson's correlation, and kappa as measures of inter-examiner reliability. Journal of Dental Research, 65(2): 128-30.
26. Hosseinzade, F. and A. Salageghe. 2013. Study and Comparison of Partitioning Clustering Algorithms. Iranian Journal of Medical Informatics, 2(1): 38-42.
27. Jafarzadeh, A., A. Mahdavi and H. Jafarzadeh. 2017. Evaluation of forest fire risk using the Apriori algorithm and fuzzy c-means clustering. Journal of forest Sscience, 63: 370-380.
28. Jahdi, R., M. Salis, A.A. Darvishsefat, F. Alcasena, M.A. Mostafavi and V. Etemad. 2016. Evaluating fire modelling systems in recent wildfires of the Golestan National Park, Iran. Forestry: An International Journal of Forest Research, 89(2): 136-49.
29. Janmenjoy, N., N. Bighnaraj and H.S. Behera. 2015. Fuzzy C-Means (FCM) Clustering Algorithm: A Decade Review from 2000 to 2014. Computational Intelligence in Data Mining, 133-49 pp.
30. Karimov, J., M. Ozbayoglu and E. Dogdu. 2015. k-Means Performance Improvements with Centroid Calculation Heuristics Both for Serial and Parallel Environments. IEEE International Congress on Big Data (BigData Congress). New York, USA, 444-451 pp.
31. Kaufman, L. and J.P. Rousseeuw. 2005. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley series in probability and mathematical statistics. Hoboken, NJ: Wiley-Interscience, 342 pp.
32. Khatami, A., S. Mirghasemi, A. Khosravi, C.P. Lim and S. Nahavandi. 2017. A new PSO-based approach to fire flame detection using K-Medoids clustering. Expert systems with applications, 68: 69-80.
33. Khatami, A., S. Mirghasemi, A. Khosravi and S. Nahavandi. 2015. An efficient hybrid algorithm for fire flame detection. International Joint Conference on Neural Networks (IJCNN). Killarney, Ireland, 1-6 pp.
34. Krishnapuram, R., A. Joshi and Y. Liyu. 1999. A fuzzy relative of the k-medoids algorithm with application to web document and snippet clustering. IEEE International Fuzzy Systems, 1281-1286 pp.
35. Lewis-Beck, M.S. and A. Skalaban. 1990. The R -Squared: Some Straight Talk. Polit. Anal. 2: 153-71.
36. Liang, M. and H. Zhang. Wang. 2019. A Neural Network Model for Wildfire Scale Prediction Using Meteorological Factors. IEEE Access, 7: 176746-55.
37. Littell, J.S., D.L. Peterson, K.L. Riley, Y. Liu and C.H. Luce. 2016. A review of the relationships between drought and forest fire in the United States. Global change biology, 22(7): 2353-69.
38. Mesakar, S. and M. Chaudhari. 2013. A Review of Clustering Algorithms, 249: 7-28.
39. Miyamoto, S., H. Ichihashi and K. Honda. 2008. Algorithms for Fuzzy Clustering - Methods in C-Means Clustering with Applications. Springer, 2008th Edition: 258 pp.
40. Mood, A.M., F.A. Graybill and D.C. Boes. 2013. Introduction to the theory of statistics. McGraw Hill, 3rd Edition, New Delhi, India, 480 pp.
41. Pham, B., A. Jaafari, M. Avand, N. Al-Ansari, T. Du, H. Phan, T.V. Phong, D.H. Nguyen, L.V. Lie, D. Mafi-Gholami, I. Prakash, H. ThiThuy and ThiTuyen. 2020. Performance Evaluation of Machine Learning Methods for Forest Fire Modeling and Prediction maps provide a basis for developing more efficient fire-fighting strategies and reorganizing policies in favor of sustainable management of forest resources. Symmetry, 12: 1-21.

42. Rahimi, I., S.N. Azeez and I.H. Ahmed. 2020. Mapping Forest-Fire Potentiality Using Remote Sensing and GIS, Case Study: Kurdistan Region-Iraq. In: Al-Quraishi AMF, AM. Negm (eds.). Environmental remote sensing and GIS in Iraq. Springer Water, 2364-6934.
43. Sadeghifar, M., A. Beheshti Alagha and M. Por Reza. 2016. Variability of Soil Nutrients and Aggregate Stability in Different Times after Fire in Zagros Forests (Case Study: Paveh Forests). Ecology of Iranian Forest, 4: 19-27 (In Persian).
44. Sekizawa, A. 2005. Fire Risk Analysis: Its Validity and Potential for Application in Fire Safety. Fire Safety Science, 8: 85-100. doi:10.3801/IAFSS.FSS.8-85.
45. Shahin, M., M. Jaksa and H. Maier. 2008. State of the Art of Artificial Neural Networks in Geotechnical Engineering. Electronic Journal of Geotechnical Engineering, 7(1): 33-44.
46. Späth, H. 1985. Cluster dissection and analysis: Theory FORTRAN programs, examples / Helmuth Späth; translator Johannes Goldschmidt. Halsted Press. New York, 226 pp.
47. Tien Bui, D., H. Van Le and N.D. Hoang. 2018. GIS-based spatial prediction of tropical forest fire danger using a new hybrid machine learning method. Ecological Informatics, 48: 104-16.
48. Velmurugan, T. 2011. A Comparative Analysis between K-Medoids and Fuzzy C-Means. Journal of Theoretical and Applied Information Technology, 27(1): 19-30.
49. Velmurugan, T. 2012. Evaluation of k-Medoids and Fuzzy C-Means clustering algorithms for clustering telecommunication data. International Conference on Emerging Trends in Science, Engineering and Technology, 115-120 pp.
50. Wackerly, D.D., W. Mendenhall and R.L. Scheaffer. 2008. Mathematical statistics with applications. 7th ed. Cengage Learning. United State, 994 pp.
51. Wei, C.P., Y.H. Lee and C.M. Hsu. 2000. Empirical Comparison of Fast Clustering Algorithms for Large Data Sets. 33rd Annual Hawaii International Conference on System Sciences, 351-363 pp.
52. Xu, R. and D. Wunsch. 2005. Survey of clustering algorithms. IEEE Transactions on neural networks, 16(3): 645-78.
53. Yassemi, S., S. Dragićević and M. Schmidt. 2008. Design and implementation of an integrated GIS-based cellular automata model to characterize forest fire behaviour. Ecological Modelling, 210(1-2): 71-84.

Comparison of the Performance of Fuzzy C-Means and K-Medoids in Modeling Forest Fire Occurrence (Case Study: Saravan Forests, Gilan)

Shaghayegh Zolghadry[1], Mehrdad GhodsKhahDaryaei[2], Kamran Nasirahmadi[3] and Esmaeil Ghajar[4]

1- Ph.D. Student, Department of Forestry, Faculty of Natural Resources, Guilan University, Iran
(Corresponding author: Shaghayegh.zolghadry@gmail.com)
2 and 4- Associate Professor and Assistant Professor, Department of Forestry, Faculty of Natural Resources, Guilan University, Iran
3- Ph.D. In Environments, Faculty of Fisheries and Environment, Gorgan University of Agricultural Sciences and Natural Resources
Received: August 16, 2020 Accepted: September 28, 2020

Abstract

Hyrcanian (Caspian) area is one of the most important vegetation areas in Iran, which due to its antiquity, has a high ecosystem value. On the other hand, this ecosystem is involved in multiple fires every year and loses a significant level of vegetation, so the use of scientific methods to predict places with potential fire risk is very important. This can be used for the conservation management of Hyrcanian forests. Many real-world systems are used in terms of pattern recognition, so proper use of machine learning methods is essential in practical applications. However, the use of clustering-based methods is emphasized as an effective method due to its approach in pattern recognition and output discovery. The purpose of this study was to evaluate the ability and compare the performance of Fuzzy C-Means and k-Medoids clustering in modeling forest fire occurrence with emphasis on the performance capabilities of the algorithm. Due to the existence of periodic fires, the mentioned algorithms were used to improve the level of coding in MATLAB software in order to improve studies in the field of forest fire risk prediction. Model input criteria in this study are recorded fire points, distance to agricultural areas, distance to the road, distance to the river, air pressure, solar radiation, slope, aspect, wind speed, forest type and percentage of canopy density. The results obtained from the fire hazard prediction map of both algorithms show their high ability to predict the fire occurrence model. Also, based on the results of the confusion matrix table of the comparison of the two algorithms, the FCM algorithm showed better performance than the k-medoids algorithm in predicting places with potential fire risk. Therefore, the use of FCM algorithm is suggested as one of the effective methods in differential clustering for future studies.

Keywords: Clustering Algorithm, Fire Occurrence Modeling, Saravan Forest

[DOR: 20.1001.1.24237140.1400.9.17.17.1]
[DOI: 10.52547/ifej.9.17.163]